N92-14093

COLORED NOISE EFFECTS ON BATCH ATTITUDE ACCURACY ESTIMATES[1]

By Stephen Bilanow
General Sciences Corporation
6100 Chevy Chase Drive
Laurel, Maryland 20707

Abstract

This paper investigates the effects of colored noise on the accuracy of batch least squares parameter estimates with applications to attitude determination cases. The standard approaches used for estimating the accuracy of a computed attitude commonly assume uncorrelated (white) measurement noise, while in actual flight experience measurement noise often contains significant time correlations and thus is "colored." For example, horizon scanner measurements from low Earth orbit have been observed to show correlations over many minutes in response to large scale atmospheric phenomena.

A general approach to the analysis of the effects of colored noise is investigated, and interpretation of the resulting equations provides insight into the effects of any particular noise color and the worst case noise coloring for any particular parameter estimate. It is shown that for certain cases, the effects of relatively short term correlations can be accommodated by a simple correction factor. The errors in the predicted accuracy assuming white noise and the reduced accuracy due to the suboptimal nature of estimators that do not take into account the noise color characteristics are discussed. The appearance of a variety of sample noise color characteristics are demonstrated through simulation, and their effects are discussed for sample estimation cases. Based on the analysis, options for dealing with the effects of colored noise are discussed.

INTRODUCTION

equirement for flight dynamics support is the estimation of the accuracy of attitude and orbit solutions, and this uires a knowledge of the measurement noise characteristics. Often, the measurement errors are assumed to be ependent and identically distributed, what engineers commonly call "white" noise. One reason this assumption nade is simply that noise of this nature is easy to handle in estimation algorithms. However, this is not always orrect assumption for real spacecraft data. This paper investigates the implications of that assumption, cusses a formulation for calculating the true parameter uncertainty when the noise is not white, and shows how interpret the effects of various noise colors in some representative cases.

olored" noise refers to any noise that is not white, i.e., that has correlations related to the time between asurements of the same type. "Batch" refers to the computation of fixed parameters using data over a given ie span in a single solution.

1 COLORED NOISE IN SPACECRAFT DATA

acecraft horizon scanner data provides a clear example of measurement noise that is obviously non-white, and r which an explanation for long term correlations of various frequencies is apparent. Figure 1 shows a sample anner data from Seasat and Landsat. In the Seasat mission, the bumps in the data were directly correlated with e infrared scanner "seeing" a high altitude cloud in the threshold adjust region of the horizon detection logic eference 1). Thus large scale atmospheric phenomena contributed a low-frequency "noise" to the scanner easurements. In the Landsat mission, the "bumps" could not be correlated with specific cloud features; however ng term correlations are clearly present (note that the highest frequency component of the Landsat data noise s filtered by 128 point averaging for data volume reduction; the remaining noise variations clearly have longer rrelations than white noise.) For Landsat some of the very long term variability was associated with seasonal

This work was supported by the NASA Small Business Innovation Research (SBIR) program

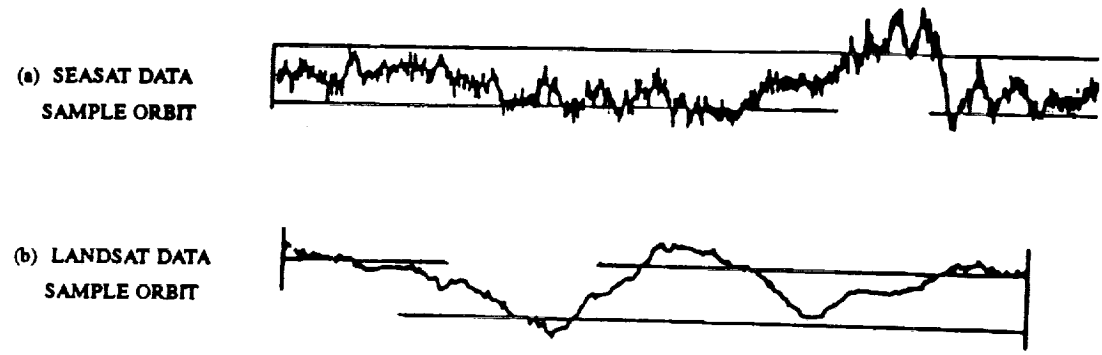


Figure 1. Sample Horizon Scanner Data from Seasat and Landsat

random stratospheric temperature variations (Reference 2). Ever since horizon scanners have first been buil manufacturers have worked to make them less sensitive to clouds in the lower atmosphere and to trigger on th more stable stratosphere, but yet it seems natural that this sensor may remain sensitive to large scale "weathe phenomena to some extent and thus have long term correlations in the data.

Other sources of long term correlations in sensor measurements can include any modeling uncertainties such a sensitivities to stray light, magnetic field changes (external or internally generated), or temperature variation Certainly the line between noise and systematic errors can become blurred, but the low frequency noise model fo some types of possible modeling errors can be useful. Another source of effective low frequency noise can b spacecraft dynamics modeling uncertainties or the effects of gyro noise. However, the similarity of the effects o these noise types with low frequency measurement noise will not be developed in this paper.

1.2 BRIEF LITERATURE REVIEW

The equations for treating colored noise in batch least squares estimation have long been known and are given i numerous textbooks. It is a matter of applying an optimal data weighting based on the expected correlations However, as a practical matter, many actual estimation systems simply assume white noise. Although ever relevant text reviews the optimal, maximum likelihood weighting, and the simplification with the white nois assumption, there is surprisingly little discussion of the impact of this simplifying assumption and what it can mea in practical batch estimation problems. Furthermore, there is a relatively simple formula for computing th accuracy of an estimator that assumes white noise while actual correlations are present. This formula does no seem to be noted, let alone its relevance emphasized, in most texts on estimation. The general form of thi equation, giving the errors due to a difference between any assumed and actual noise covariance, is given in th mathematics for the general model for attitude determination error analysis developed at GSFC (Reference 3). However, in the current system implementation based on this analysis, only white noise assumptions are allowed (Reference 4). References 5 and 6 both mention this formula and discuss the implications briefly by an example. It is likely that more attention to this problem may be contained in the broad literature on estimation in various fields, but its consideration (particularly for flight dynamics applications) seems to be very infrequent.

There is notable available literature on handling colored noise in Kalman filter applications. Problems in handling colored noise in continuous time filters were first presented and resolved by Bryson and Johansen in 1965 (Reference 7), and further developments were provided and a few practical applications were discussed in papers that followed (References 8 through 12). Sections on handling "colored noise" in Kalman filters are found in books on estimation (e.g., References 5, and 13 through 15) published in the early 70's. These references give prescriptions for optimally filtering the data given colored noise. However, these references do not generally address a sensitivity analysis to the "suboptimal" white noise assumption in covariance analysis, which is the problem discussed in this paper in the batch estimation case.

mation and the spectral analysis of data in general is a field with a long history, wide application, and siderable development. Today, voluminous literature on spectral analysis and estimation is found in munications engineering, statistical time series analysis, time standards stability, and speech data processing, ng other fields. Although effort was made to locate relevant references, an exhaustive survey is by no means ned.

MATHEMATICAL DEVELOPMENT

OPTIMAL WEIGHTED LEAST SQUARES

assume a linearized model of our measurements, z,

$$z = Hx + \epsilon \tag{1}$$

re H is the matrix of partials of the measurements with respect to the state parameters, x is the state vector of meters, and ϵ is the vector of measurement errors. Basic weighted batch least squares provides an estimate of state parameters, $\hat{x}$,

$$\hat{x}_o = (H^T W H)^{-1} H^T W z \tag{2}$$

; estimate is optimal if the weight matrix is the inverse of the measurement noise covariance matrix

$$W = R^{-1} \tag{3}$$

re R is the expected noise covariance

$$R = E\left[\epsilon\ \epsilon^T\right] \tag{4}$$

accuracy of an estimate is given by the state covariance matrix

$$P_o = (H^T W H)^{-1} \tag{5}$$

se equations are the maximum likelihood estimate, or best linear unbiased estimate, and they are equivalent to Bayesian estimate if no a priori uncertainty information is available.

WHITE NOISE/UNWEIGHTED LEAST SQUARES

ve know that our measurements are independent and uncorrelated, then R is a diagonal matrix. If we make the itional assumption that all the measurements have the same variance, then we may write R as a scalar times the ntity matrix, I.

$$R = \sigma_n^2 I \tag{6}$$

$$W = \frac{1}{\sigma_n^2} I \tag{7}$$

his case the estimator (2) simplifies to

$$\hat{x} = (H^T H)^{-1} H^T z \tag{8}$$

l the covariance of our estimate is given by

$$P_w = \sigma_n^2 (H^T H)^{-1} \tag{9}$$

This simplification is commonly made in many systems, including those for attitude determination. Two reaso often given are: (1) A priori measurement statistics may not be available, and it is often just assumed th independence of the measurements is a good model, and (2) It is also sometimes observed that optimal weighti requires computing the inverse of the measurement noise covariance, and this may not be practical when handli large amounts of data. This is an additional motivation for assuming equal weighting is good enough. (It is n widely noted that one basic colored noise model does have an exact form for its inverse -- see Equation 23.)

It is the purpose of this paper to investigate the impact of this simplification on the expected covariance of o estimate. This can be particularly important for prelaunch studies when we want to predict how well our estimat will perform. It also can be of importance for postlaunch analysis if we want to use the estimators predict covariance as an indicator of the actual attitude accuracy.

2.3 *UNWEIGHTED ESTIMATOR WITH COLORED NOISE*

The expected variance of the unweighted least squares estimator in the presence of correlated measurements may b derived directly by taking the expected covariance of estimator (8) assuming noise covariance R.

$$P_s = (H^T H)^{-1} H^T R H (H^T H)^{-1} \qquad (9)$$

Thus if we have a model for the actual noise covariance, R, we can directly compute the error of our unweighte estimator. This is the main formula used to derive results presented in this paper. As observed in the literatur review, it is remarkable how seldom this equation is noted.

As we shall see, interpreting results from this formula requires some careful attention. Note that there are as man terms in the noise covariance as there are points being fit in the least squares estimation. This gives a tremendou amount of power in terms of possible assumptions about our noise model. For example, this formula can be use to evaluate the effects of random biases as well as noise in the traditional sense.

In the terminology often used in error analysis, the unweighted least squares is considered a suboptimal estimator i the context that actual correlations are present in the noise (and hence the choice of subscript). Note however tha we are not primarily concerned here with the actual performance of this suboptimal estimator relative to the optima one, although we will make observations about this difference ($P_s - P_o$). Instead we will be concerned mainly abou the erroneous prediction of the suboptimal estimator accuracy assuming white noise relative to its actual accurac given colored noise ($P_s - P_w$). As we shall see, this suboptimal estimator does not generally do badly relative to th optimal one, but the prediction of its accuracy erroneously assuming white noise can be quite unrealistic.

It is noted that a more general equation for error analysis can be obtained by taking the expected covariance for the weighted least squares estimator when the true noise covariance is different than the expected noise covariance. We will not, however, look at that more general problem here.

2.4 *CORRECTION FACTOR INTERPRETATION*

A very interesting and elegant interpretation can be made of the correction factor between white noise predicted accuracy and accuracy in colored noise. We take Equation (10) and break it into two parts, one giving the white noise predicted covariance, P_w, and a correction matrix C, so that

$$P_s = \left[\sigma_m^2 (H^T H)^{-1}\right] \cdot \left[\sigma_m^{-2} H^T R H (H^T H)^{-1}\right] = P_w C \qquad (1$$

)r simplicity, consider estimating a single parameter from a single time series of correlated measurements. Let 'be the true variance of the measurements and $\rho(k)$ be the autocorrelation as where k gives the sample lag. ρ is)rmalized so $\rho_o = 1$, and $\rho(k) = \rho(-k)$ due to the properties of the autocorrelation function (see Section 3 for)re discussion of noise properties). The measurement noise covariance is

$$R = \sigma_T^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_N \\ \rho_1 & 1 & & \\ \vdots & & \ddots & \\ \rho_N & & & 1 \end{bmatrix} \tag{12}$$

)d the partials matrix H is now a vector which we shall call h and refer to as our "basis vector" since it is the nction we are fitting in a least squares sense. Thus our scalar correction factor is

$$c = \frac{1}{\sigma_m^2} h^T R h \, (h^T h)^{-1} \tag{13}$$

Written out as a summation, we can write

$$c = \frac{\sigma_T^2}{\sigma_m^2} \sum_{i=-N}^{N} \rho_i \frac{\sum_{j=0}^{N-i} h_j h_{j+i}}{\sum_{j=0}^{N} h_j^2} \tag{14}$$

1 this form, the inner sum in the numerator may be recognized as the convolution of the basis vector with its)flection, or equivalently as the autocovariance function for the basis vector. The sum shown in the denominator)rmalizes the basis vector autocovariance to unity at zero lag, and thus this whole expression may be considered); a "basis vector autocorrelation" sequence, which we will label η. The correction factor is the ratio of the actual) expected variance times the dot product or projection of two normalized sequences: the true noise utocorrelation, ρ, and the basis vector autocorrelation, η.

$$c = \frac{\sigma_T^2}{\sigma_m^2} \left(\vec{\rho} \cdot \vec{\eta} \right) \tag{15}$$

The ratio of variances is just a correction for the assumed and true noise variance. If we had assumed the correct 'ariance, but had ignored the correlations at non-zero lags, the correction would be just the indicated projection.

This projection may be interpreted in the frequency domain as well. Using Parseval's theorem as applied to finite eries, the product of terms in the time domain is related to the product in the frequency domain. This is a special ase of the fact that the product in the time domain is a convolution in the frequency domain, but where we are oncerned the "DC" component in the time domain which is given by the spectrum evaluated at zero frequency. Let the Discrete Fourier Transform be defined as

$$DFT(f_n) = \sum_{h=0}^{N-1} f_n e^{-j\frac{2\pi n k}{N}} \tag{16}$$

Using Parseval's theorem gives

$$\vec{\rho} \cdot \vec{\eta} = \frac{1}{N} \overrightarrow{DFT(\rho)} \cdot \overrightarrow{DFT(\eta)} \tag{17}$$

The transform of the autocorrelation is the power spectrum. Thus the correction factor is related to the projecti of the true noise power spectrum and what we may define as the "basis vector power spectrum."

The easiest way to apply this interpretation to a multiple parameter case is to choose a set of orthogonal bas vectors so that P_w, C, and P_s are all diagonal matrices and the parameter estimates can be decoupled. The use this interpretation of the correction factor will be discussed with some specific examples in Sections 4 and 5.

2.5 LINEAR COMBINATIONS OF COLORS

At times it will be useful to consider the noise covariance as the combination of two different noise types. In th case, since Equation (10) is linear in **R**, we have

$$P_s = (H^T H)^{-1} H^T (R_1 + R_2) H (H^T H)^{-1}$$

$$= (H^T H)^{-1} H^T R_1 H (H^T H)^{-1} + (H^T H)^{-1} H^T R_2 H (H^T H)^{-1} \quad ($$

Thus if the effects of two independent noise sources are evaluated, the total effect from both can be computed the linear sum of the variances due to their separate effects. (Note that the combination is linear in the varianc not linear in the standard deviation of the noise.)

3. COLORED NOISE SAMPLES

This section defines some specific types of colored noise for analysis and provides examples for illustration.

3.1 NOISE SIMULATION AND CHARACTERIZATION

Stationary noise of any desired spectrum can be obtained by passing white noise through an appropriate filter. An stable time invariant linear filter will color a white noise input according to its frequency response. Since ther are as many possible "colors" to noise as there are frequency response curves, which is an uncountable infinity c curves, we will necessarily restrict our attention to a few simple classes of coloring for illustrating specific case The theory of digital filtering and time series analysis is covered in numerous texts (e.g. Ref. 16-19). For thi discussion we will simply provide a few definitions to clarify the noise models that will be used in the sampl cases that follow.

In the time domain, a linear filter is defined by its impulse response which when convolved with its input, in ou case white noise, produces the system output, colored noise. The variance of the output noise from a filter will b given by the sum of squares of the impulse response sequence. In the examples shown we will routinel normalize the output variance to unity and have the plot scales cover +/- 3 standard deviations for consistency.

The most efficient way to generate colored noise for fairly simple processes is through linear difference equations Care must be given to the initial conditions in the noise generation to assure immediately stationary realization in statistical sense (Reference 19), otherwise the noise must be simulated for a period to reach a steady stat (particularly for long lag process simulations).

A stationary stochastic (noise) process is characterized by its autocovariance function or alternatively by the Fourie transform of the autocovariance function which is its power spectral density (PSD). The autocovariance is define as

$$\gamma(k) = E\left[x(n)\ x(n+k) \right] \quad (1$$

will refer to the autocorrelation function (ACF), which is the autocovariance normalized by its value at zero (which is its variance),

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \tag{20}$$

ce this is the noise characterization that enters directly into our formulas for evaluation of the colored noise ects on a least squares estimation accuracy, we note the autocorrelation functions for the sample noise cesses presented below.

WHITE NOISE

ure 2 shows the appearance of white noise with uniform and Gaussian probability distributions, both of which familiar to those with data processing experience. As stated earlier our definition of stationary white noise is : that the data samples are independent and identically distributed. The most common assumption about the ribution is that it is Gaussian because of the tractable statistical properties. We will use the Gaussian noise as ut to filters to simulate the colored noise shown here, but it is noted that the choice of uniform or Gaussian ut distributions does not noticeably affect the appearance of the filtered noise. A result of the central limit orem is that the more heavily filtered the noise is, the more the output distribution will approach Gaussian no ter what the input distribution.

te also that the number of data samples plotted and the plot scaling impacts the visual appearance of any noise. use 400 points for each of the plots shown here for uniformity. The plot scales are set at the expected value three standard deviations. Data quantization can also significantly impact the appearance, but we will not ulate quantization here.

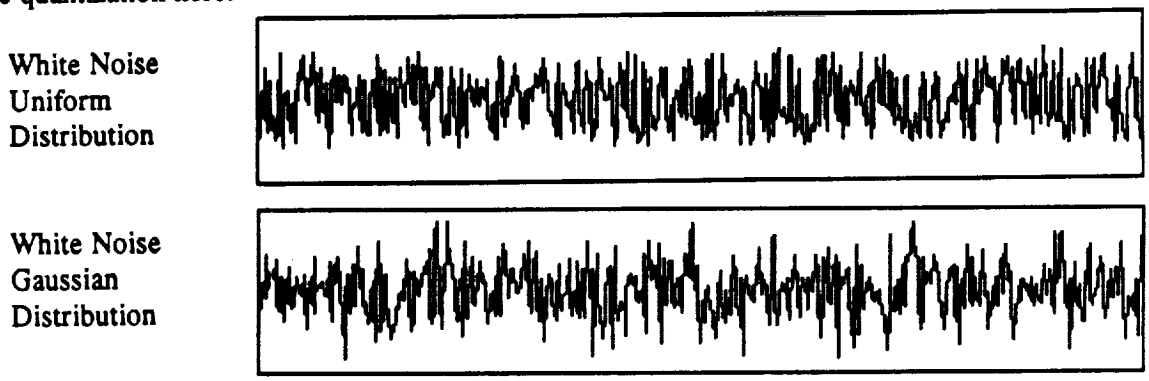


Figure 2. Sample White Noise with Uniform and Gaussian Distributions

LOWPASS NOISE

simple single pole lowpass filter of white noise, w(n), is specified by the linear difference equation:

$$x(n) = \phi\, x(n-1) + w(n) \tag{21}$$

ere ϕ is the pole location. This is known as a first order autoregressive process (AR(1)--a label we will use for evity). It is also commonly called a first order stationary Markov process. The autocorrelation for this process given by

$$\rho(k) = \phi^{|k|} \tag{22}$$

Thus ϕ gives the correlation between consecutive samples. Samples of this type of noise for selected values of th ϕ are shown in Figure 3. The values of ϕ shown correspond to effective "time constants - τ" (for the correlatio to decay to 1/e) of 2, 15, and 100 samples duration.

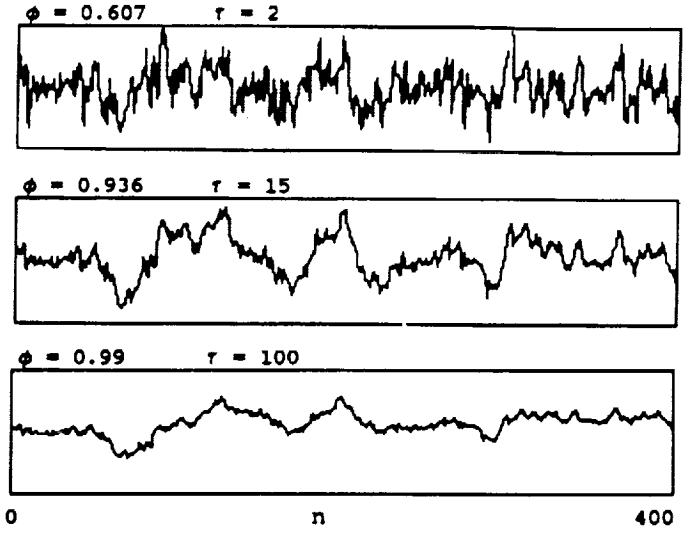


Figure 3. Sample Lowpass Noise with Selected Time Constants (400 Samples Plotted)

Since the general difficulty of inverting the noise covariance matrix R is sometimes cited as a reason for not attempting th optimal weighting, it is interesting to note that for this particular noise model, the noise covariance matrix has an exact inver

Noise Covariance

$$R = \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^N \\ \phi & 1 & \phi & & \\ \phi^2 & \phi & 1 & & \\ \vdots & & & \ddots & \phi \\ \phi^N & \cdots & & \phi & 1 \end{bmatrix}$$

Optimal Weighting

$$W = R^{-1} = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & -\phi & & & 0 \\ -\phi & 1+\phi^2 & -\phi & & \\ & -\phi & 1+\phi^2 & \ddots & \\ & & \ddots & 1+\phi^2 & -\phi \\ 0 & & & -\phi & 1 \end{bmatrix} \quad (2$$

Another type of lowpass filter is a simple N-point running average filter. The autocorrelatior function is a finite triangular shaped sequence:

$$\rho(k) = 1 - \frac{|k|}{N}, \quad |k| < N, \ 0 \text{ otherwise} \quad (2$$

A sample 23-point running average filter of the same input white noise sequence is shown in Figure 4. Note the similarity with the AR(1) process with ϕ = .936. This similarity was emphasized by choosing the number of points so that the above finite autocorrelation function was a simple linear approximation to the AR(1) exponential decay curve. This illustrates how the appearance of many of the general visual features in the filtered noise are the same for filters with basically the same short term correlations. The long tail in the AR(1) process does not significantly influence the visual appearance of the noise.

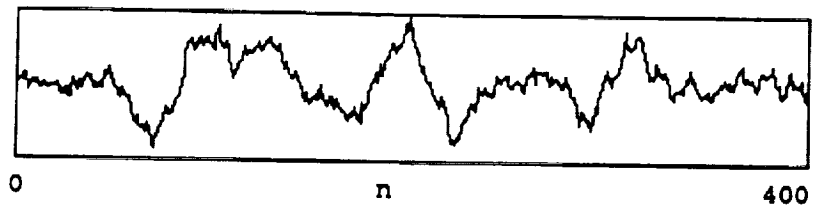


Figure 4. Sample 23-point Running Average Filter of White Noise

HIGHPASS NOISE

e AR(1) lowpass filter becomes a simple highpass filter for ϕ less than zero. A sample highpass noise plot is shown in ure 5.

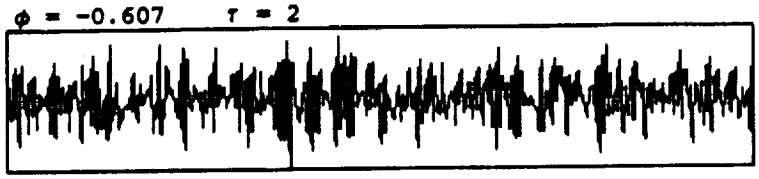


Figure 5. Sample Highpass Noise, ϕ = -0.607

BANDPASS NOISE

generate noise with a selected frequency emphasis, we will utilize a simple two pole filter with complex njugate roots so that our impulse response function remains real. This noise process is a second order oregressive AR(2) model or second order stationary Markov process. In terms of pole locations at radius r and gle Θ around the unit circle, the linear difference equations for generating this noise are given by

$$x(n) = w(n) + 2r\cos\theta\, x(n-1) - r^2 x(n-2) \tag{25}$$

e autocorrelation function is given by

$$\rho(k) = r^k \left[\cos(k\theta) + \frac{\cos\theta}{\sin\theta} \frac{1-r^2}{1+r^2} \sin(k\theta) \right] \tag{26}$$

o samples of noise generated in this way are shown in Figure 6 for a relatively high and relatively low quency emphasis.

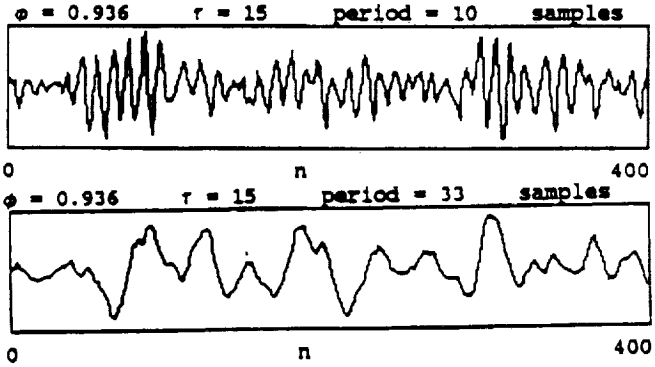


Figure 6. Bandpass Filter Noise with Two Different Frequencies Emphasized

COMBINED NOISE MODELS

ises of any particular types can be combined and it is important to note that a low amplitude of one "color" can hidden by the dominance of another, although it seems that human eye and brain do a pretty good job of scriminating patterns. For example, Figure 7 shows a combination of independent white noise of standard viation 0.8 with the moderate lag lowpass noise, AR(1), ϕ equal to 0.936, of standard deviation 0.6 (the total riance is $(0.6)^2 + (0.8)^2 = 1.0$). The total effect on estimation accuracy will equal the combination of their parate effects as noted in Section 2.5.

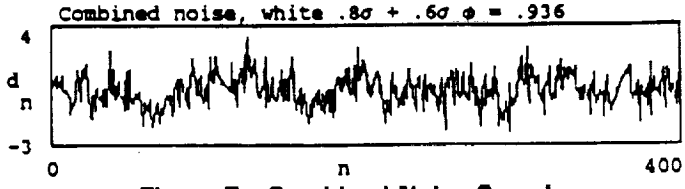


Figure 7. Combined Noise Sample

4. ESTIMATION OF THE MEAN

It is instructive to start with the simplest of cases in order to understand the effects of colored noise on estimation. Thus begin by reviewing the effects of colored noise in the estimation of the mean of a single sequence of measuremen Although this is a simple case it can be considered as basically applicable to some attitude estimation cases, for example i spacecraft is inertially pointing and collecting star measurements under basically the same geometry. Further, when seve parameters are being estimated, one of the parameters, a measurement bias for example, may essentially be computed fr the mean of the measurements.

4.1 ACCURACY vs SAMPLES IN LOWPASS NOISE

We will start with the simplest lowpass filter noise, a first order autoregressive process (AR(1)) or simple Markov process defined in Equation 21. The partials of all measurements w.r.t the mean is 1, so the basis vector contains all 1's, and t unweighted estimate of the mean is the sample average. One can derive a formula for the uncertainty in the average as estimator of the mean directly through slightly tedious algebra and recognition of the proper series summations. One obtai

$$\sigma^2_{AVG} = \frac{1}{N}\left\{\frac{1+\phi}{1-\phi} - \frac{2\phi}{N}\,\frac{1-\phi^N}{(1-\phi)^2}\right\}$$

One can also compute the optimally weighted (or maximum likelihood) estimate of the mean, using the exact inverse noted Equation (23) to obtain:

$$\sigma^2_{OPT} = \frac{1}{N}\left\{\frac{1+\phi}{1-\phi+\frac{2\phi}{N}}\right\}$$

Results of the uncertainty in these various estimates of the mean are shown in Figure 8. Two different values for t correlation between samples are illustrated. Both the unweighted and weighted (suboptimal and optimal) estimates of t mean are less accurate in the lowpass noise. It is interesting to note that the unweighted estimate of the mean is almost accurate as the optimally weighted estimate even when the correlation between samples is fairly high. (The relati weighting of data points is give by the sum of the columns in the weight matrix (see Equation 23), so it is interesting to nc that the optimal weighting for this noise model simply adds more weight to the end points. One interpretation of this is th the end points carry more information because of correlations with the data beyond the end points.) On the other hand, t white noise estimate of the accuracy is unrealistically optimistic when significant lowpass noise is present.

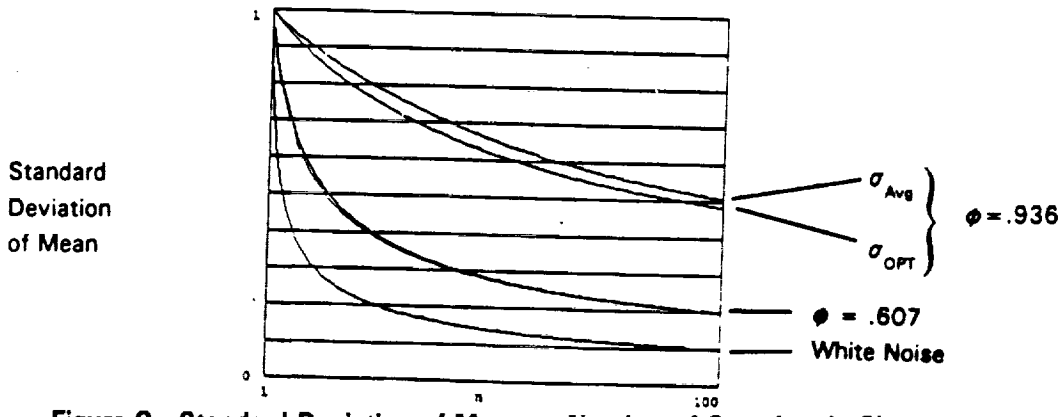


Figure 8. Standard Deviation of Mean vs Number of Samples, in Simple Lowpass Noise

4.2 ASYMPTOTIC RESULTS

A feature to note in Figure 8 is that the ratio of actual accuracy to that predicted by white noise appears consistent as th number of samples gets large. In fact, it can be seen from the formulas (27) and (28) that ratio of the accuracy of both th estimators to the accuracy assuming white noise converges to a limit for large N, which is given by:

$$\lim_{N\to\infty} \frac{\sigma^2_{AVG}}{\sigma^2_w} = \lim_{N\to\infty} \frac{\sigma^2_{OPT}}{\sigma^2_w} = \frac{1+\phi}{1-\phi} \tag{29}$$

convergence ratio, which applies to both the unweighted and weighted least squares estimates, is shown in Figure 9 as a tion of the correlation between samples, ϕ. Here the range of ϕ is allowed to go from -1 to 1 to illustrate effects from eme highpass to extreme lowpass noise. This illustrates how the accuracy in the estimate of the mean is lower in lowpass

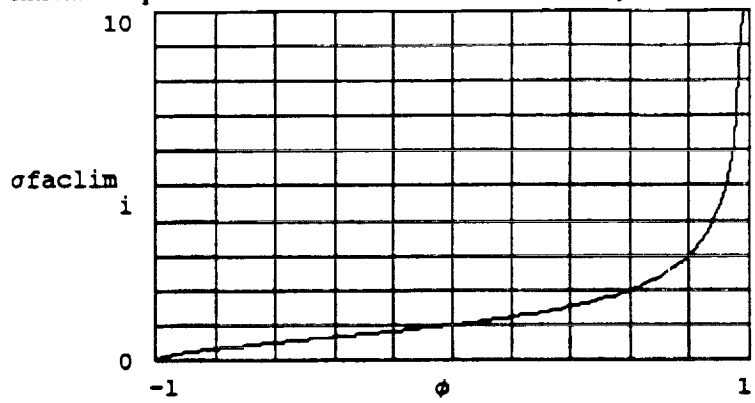


Figure 9. Asymptotic Ratio for Colored vs White Noise Accuracy

e, but higher in highpass noise. In the extreme case for highpass noise, the data may be oscillating back and h, but the expected value of the midpoint is nevertheless exactly the mean. The extreme case for lowpass noise random bias which we will discuss more later. In this case, the ratio goes to infinity because the white noise uracy converges to zero. We will later see that for certain well behaved general multiparameter cases this vergence ratio will apply approximately to all the parameters.

PROJECTION INTERPRETATION

v let us take a first look at the correction factor interpretation previously discussed as it applies in this case. We examine it in the time domain and make a brief note about the corresponding results in the frequency domain. ure 10 illustrates the noise autocorrelation for this process and the basis vector autocorrelation for a short, lium and long data span. The basis function is a constant, the convolution with itself makes the autocorrelation iangular pulse that is stretched out for longer data spans. Underneath each of the basis autocorrelation vectors is product whose sum gives us the correction factor relative to the white noise accuracy. As the data span goes to nity, the correction factor converges to the sum of the noise autocorrelation values which is a convergent metric series.

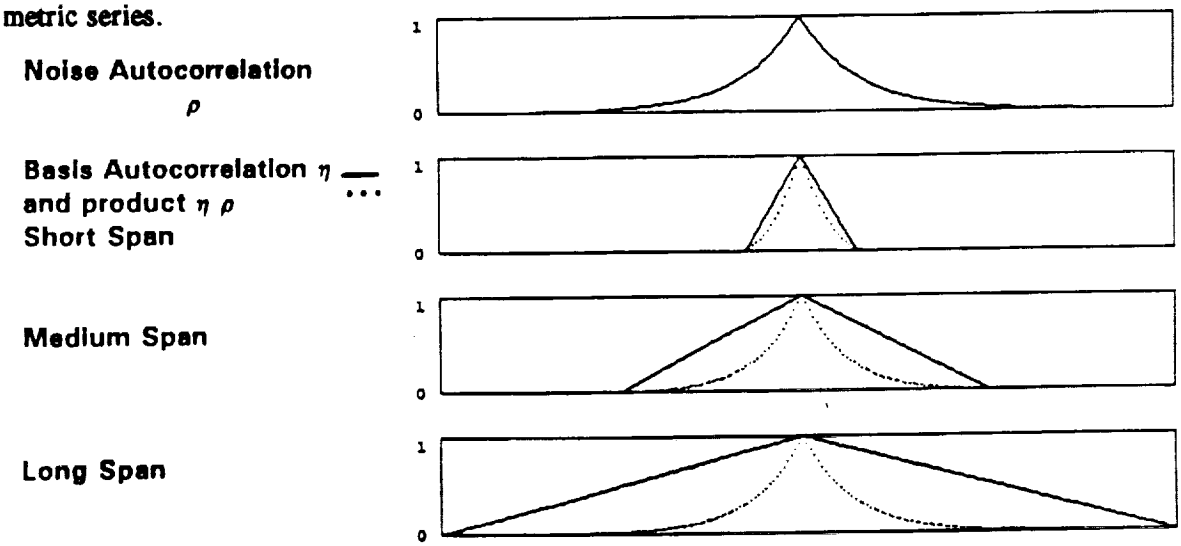


Figure 10. Projection Interpretation for Correction Factor to Estimate of the Mean

It is easy to see from geometric arguments presented in this case that, more generally, the correction factor for estimation of the mean will converge for any finite autocorrelation (MA or FIR process), or any process wh the sum of the autocorrelation terms is finite. This sum is in fact finite for any ARMA stationary process. proof of this asymptotic convergence ratio for the estimation of the mean is given in Reference 19 (Chapter 7).

4.4 INCREASING SAMPLES IN A FIXED DATA SPAN

Another aspect of the difference between white noise and colored noise is illustrated by considering an increas number of sample points taken over a fixed data span. Under the ideal white noise model, no matter how clos time the samples are taken they are still independent, so the variance decreases as the inverse of the numbe samples. In actual practice however, one expects that as samples become very close in time, they become hig dependent so that at some number of samples little additional accuracy can be obtained.

Figure 11 illustrates this for the sampling of an AR(1) process to estimate the mean. As the time between samp decreases, the correlations increase. The correlation as a function of time for the AR(1) process is modeled as exponential. Let τ be a time constant for the process, so the correlation between consecutive samples in a d span of length T divided into N samples is given by

$$\phi(N) = e^{-T/\tau_N}$$

Putting this expression for ϕ in our formula for the variance of the average and taking the limit as N goes infinity, we obtain:

$$\lim_{N\to\infty} \sigma^2_{AVG} = \frac{2\tau}{T} - \frac{2\left[1-e^{-T/\tau}\right]\tau^2}{T^2}$$

The limit for the optimally weighted estimate is

$$\lim_{N\to\infty} \sigma^2_{OPT} = \frac{2\tau}{T}$$

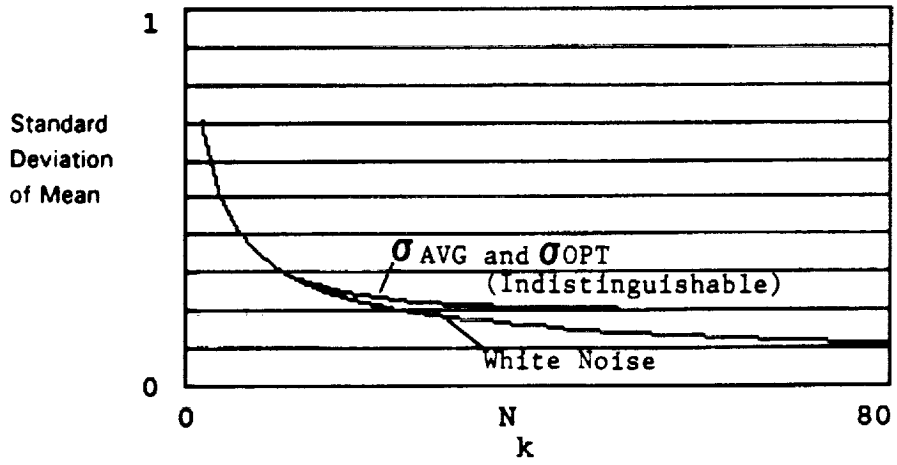


Figure 11. Increasing the Number of Samples in a Fixed Data Span

5. SPIN AXIS ESTIMATION

We now apply the analysis to a case of estimating the spin axis attitude from a single data span of r measurements which may be from a horizon scanner. We will assume a simple geometry for the problem to perr easier understanding of the results. The general nature of the results described can, however, be applied t

riety of similar attitude estimation scenarios. For example, it is similar to the computation of roll and yaw for an rth pointing spacecraft with calibrated gyro data.

GEOMETRY FOR SAMPLE CASES

e geometry for our sample cases is shown in Figure 12. We will assume a circular orbit and have the satellite in axis pointed at orbit normal, which idealizes a common mission geometry. To use round numbers (but thout loss of generality), we assume a 100 minute period orbit, so that a data span of 10 minutes, is one tenth of orbit. In order to apply convenient labels to the attitude, we will assume a polar orbit, so right ascension and clination define the spin axis in the equatorial plane without any high declination scaling concerns.

is convenient for interpretation to choose orthogonal axes for the attitude state parameters which are oriented so it there is no coupling of the errors. This axis selection to decouple the parameters can be done in any least uares estimate. For our sample cases will make those axes correspond to right ascension and declination (labeled A and DEC), by choosing our data span so that it is symmetric about the north pole point in the orbit. Thus the jor axis of the error ellipse for the spin axis will always be in the RA direction and the minor axis will be in the EC direction. To achieve generality for the orbit position one can read, instead of "RA" and "DEC," "the axis of eatest uncertainty", and "the axis of least uncertainty," respectively.

sed on this geometry, the matrix of partials of the roll measurements with respect to RA and DEC state rameters is simply a sine and cosine function of the orbit angle relative to the middle of the data span at the North le.

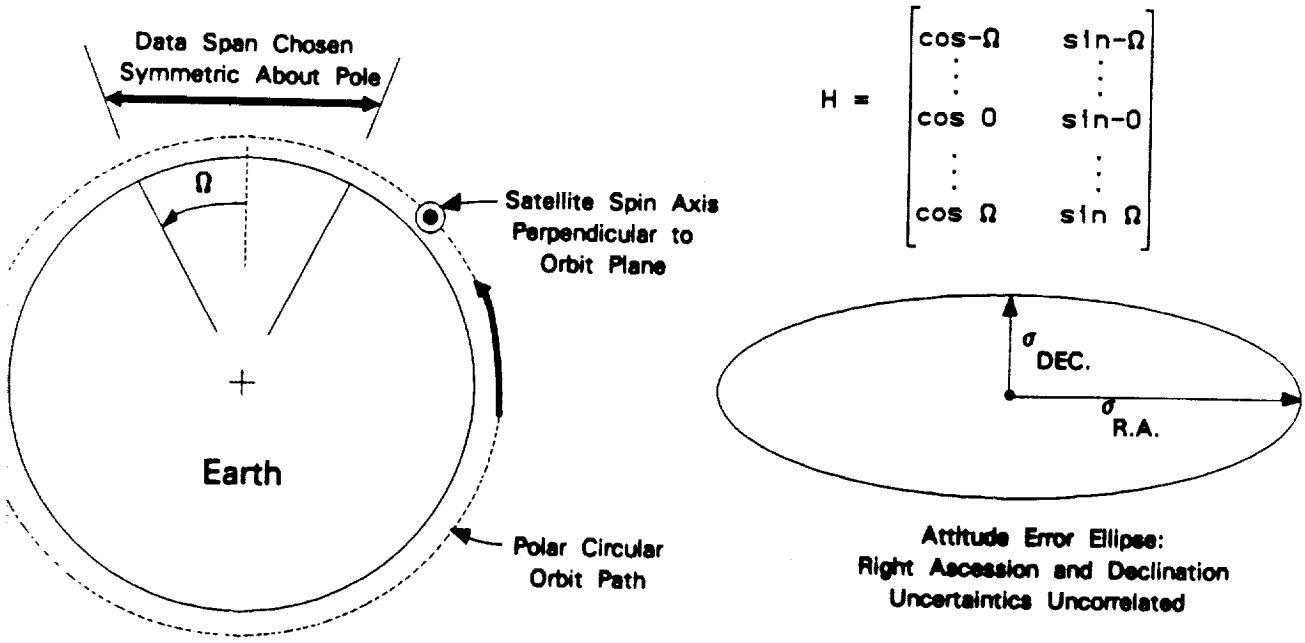


Figure 12. Geometry for Spin Axis Estimation Sample Cases

SPIN AXIS ACCURACY VERSUS TIME IN LOWPASS NOISE

gure 13 shows the DEC and RA accuracy versus time for 100 samples taken over ten minutes (1/10 orbit) here the correlation between consecutive samples is 0.607 (see Figure 7 for noise sample plot.) This corresponds a 12 second time constant on the lowpass noise. The accuracy predicted in white noise is shown for mparison, and also shown is the optimally weighted estimator accuracy which is hardly different from the weighted estimator accuracy in this case.

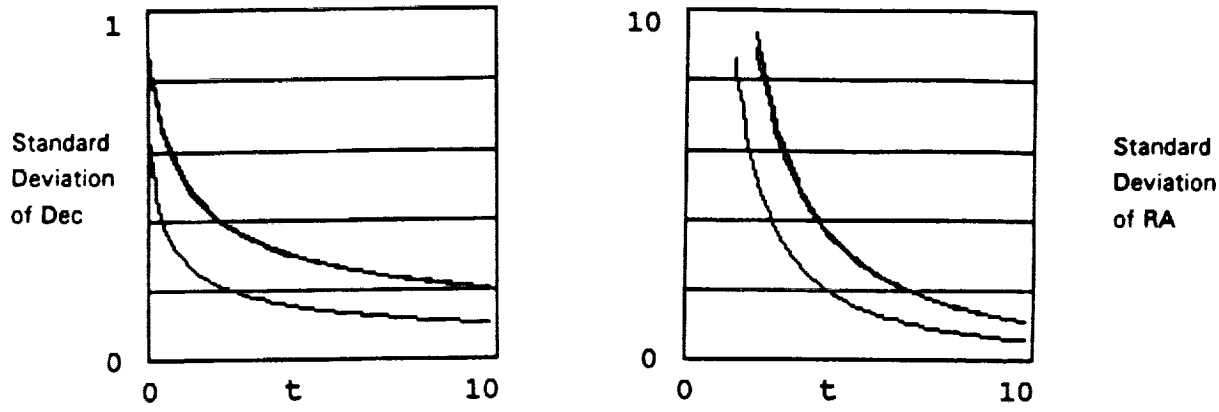


Figure 13. DEC and RA Accuracy Versus Time Over 10 Minutes

Notice that the DEC accuracy decreases in nearly exactly the same manner as the estimate for the mean illustrated for the same correlation between samples in Figure 8. This is not surprising since the basis vector for the DEC over this span, a small piece of a cosine wave, is very much like a constant.

The RA accuracy improves with the increasing data span as expected from the improved geometry that makes RA observable. Notice that the correction factor that applies to DEC estimates applies practically just as well to the RA estimates in this case.

Figure 14 illustrates the equivalent results for an extreme lowpass noise case (see Figure 7 for sample of noise). Here, something very interesting happens to the RA accuracy at very short data spans, where it is better than the accuracy predicted in white noise. An interpretation of what is happening in this case shows how the lowpass noise actually does provide better RA information. For a short data span the RA information is essentially acquired from the slope which is fit to series of observations, since the RA basis vector is a small piece of a sine wave. When the noise is highly filtered, a little piece of the data actually carries more reliable information about the slope than a group of completely random white noise measurements. In the limiting case where $\phi = 1$, the data has a random bias, but a sequence of points still retains the proper slope which will be fit properly in a least squares procedure. This limiting case is discussed further below.

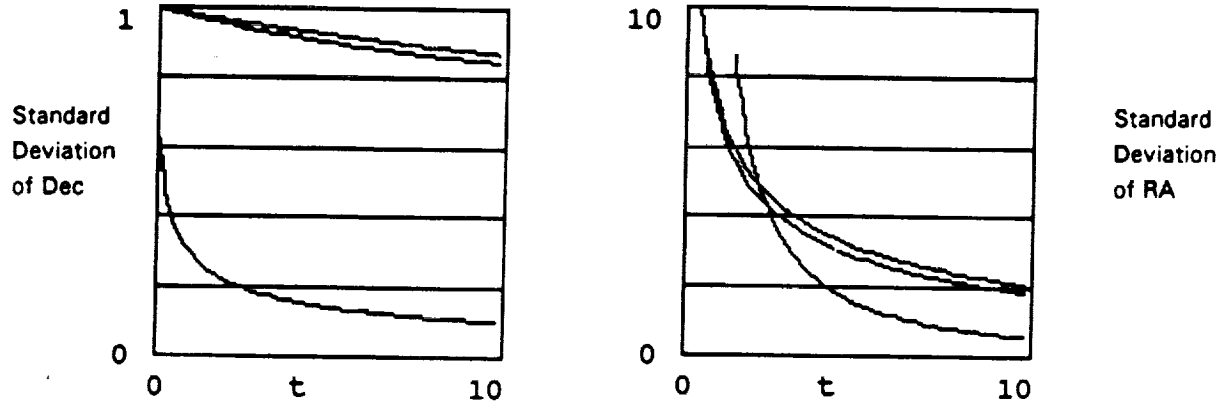


Figure 14. DEC and RA Accuracy for Extreme Lowpass Noise Case, $\phi = .99$

5.3 EFFECT OF A RANDOM BIAS ON ACCURACY

In the limiting case where the correlation term $\phi = 1$, the noise model provides the effect of a random bias (a bias that is random for each data span). To the first order, a bias affects DEC by exactly the size of the bias, but does not impact RA at all. Thus Equation 10 gives exactly this result.

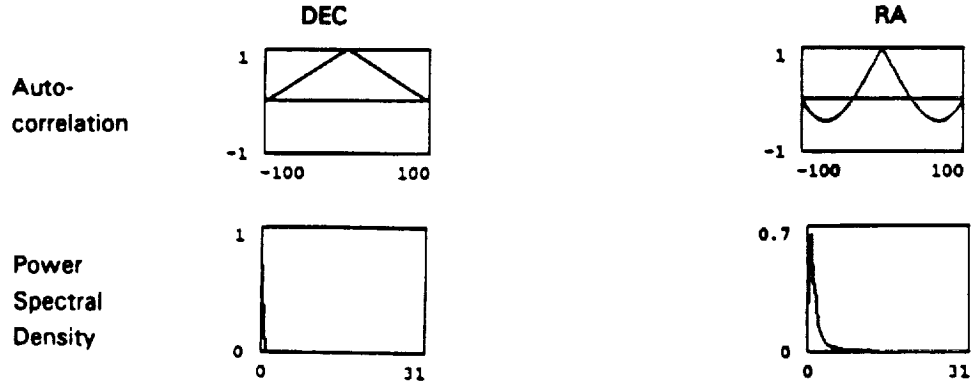
ce Equation 10 is valid for linear combinations of noise types, it is noteworthy that one can include a bias ertainty along with any colored noise model for computing the estimation accuracy. An illustration of this is limiting case of a combination of two AR(1) noise processes: one with a very short lag and one with a very g lag. In the limit, one may consider this as white noise plus a constant correlation term which is effectively a term. If one normalizes the overall noise autocorrelation function to unity with this model, one will find that RA accuracy actually improves relative to the white noise case, but it is important to recognize that one is just ctively using a smaller white noise component along with the bias component which doesn't impact the RA uracy at all. If one is careful to scale for a unity white noise component along with a bias term, the RA uracy will improve exactly as without the bias, while the DEC accuracy, which is sensitive to the bias, will rove with more observations but reach a limiting accuracy at the bias term. This result makes sense because a / long term correlation must be expected to be exactly like a bias for a finite data span.

s highlights the point that whatever noise spectrum may be worst for one parameter will not be worst for all parameters. A very long term lag is worst for estimation of the mean, and is worst for the DEC estimation in latively short data span as discussed above, but it is certainly not the worst effect on RA. Furthermore since is the most uncertain axis for this data span, long lags do not give the worst type of noise impact on the all spin axis accuracy. We will discuss the type of noise spectrum that can be worst on the overall accuracy, it will be helpful to do that after we review the insights that can be gained from our projection interpretation.

PROJECTION INTERPRETATIONS

ure 15 shows the basis vector autocorrelation and basis vector power spectral densities for the RA and DEC in 10 minute data span. The basis function for DEC, a small piece of a cosine wave, is very much like a stant, so the autocorrelation looks much like that for estimation of the mean as shown in Figure 9. The basis or power spectral density (literally the discrete Fourier transform of the sampled autocorrelation) is practically onecker delta function. The basis function for RA, a small piece of a sine wave like a linear constant slope , gives the "mustache shaped" autocorrelation shown. The power spectral density is zero at the zero uency, indicating the zero mean of the autocorrelation, and shows a peak at the lowest sampling frequency of Discrete Fourier Transform, and falls off rapidly with higher frequency. (Note the sampling frequencies of DFT correspond to sine waves with integer numbers of cycles of the data period). The DFT highlights the ntially low frequency content of these basis functions.

can see how any fairly short period correlation would cause similar effects in RA and DEC to the correction or to the white noise effects. Note that white noise is a delta function in the time domain and a constant in the uency domain. Thus a slightly broader noise autocorrelation in the time domain makes a correction factor itly greater than one.



gure 15. Autocorrelation and Power Spectral Density for DEC and RA Basis Vectors for 10 Minute Span

It is easy to see the expected effect of random bias on RA and DEC using the projection interpretation. Note th a random bias autocorrelation function corresponds to constant value of 1 while the PSD is a Kronecker del function (times N). Thus the bias has no effect on RA, while having maximum effect on DEC.

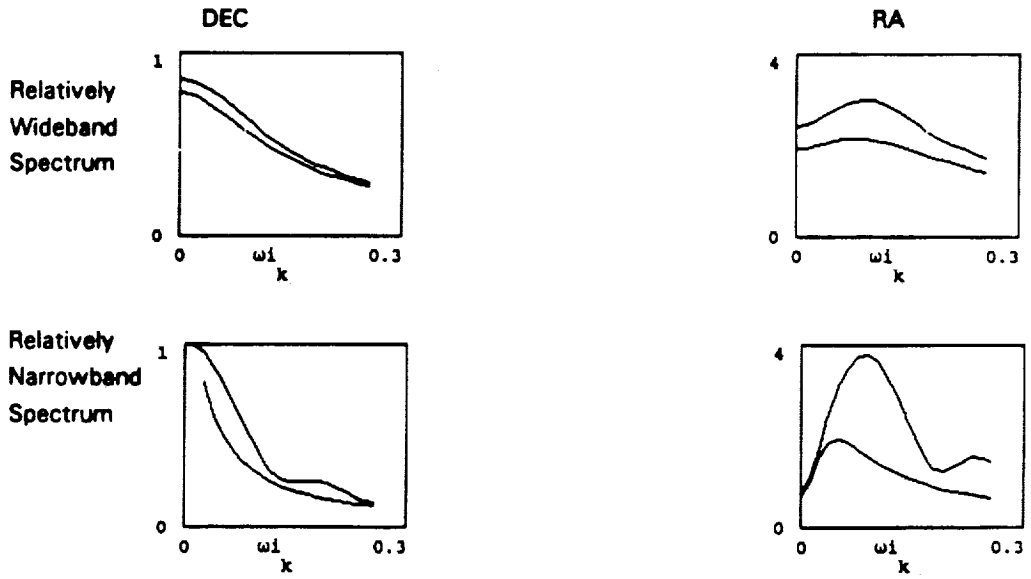
One can also use the projection interpretation to develop a sense of the worst type of noise to impact a paramet In general, one can select a noise model that has similar frequency content as the basis vector to maximum erro An extreme worst case might be a sine wave of exactly the dominant frequency of the basis. The autocorrelati function for a sine wave of random phase is a cosine function of the same frequency. In particular for RA in th case one can note by inspection of the basis autocorrelation that the worst frequency would have a period of abo 2/3 of the data span length (it would change sign at the same point as the RA basis autocorrelation).

5.5 UNCERTAINTY VERSUS NOISE COLOR

We will apply the noise model generated by a simple 2 pole filter in order to show the sensitivity of our paramet estimates to the frequency emphasis of the noise. We choose complex conjugate roots to define a real impul response. The closeness of the poles to the unit circle roughly defines the narrowness of the passband, so we w keep this distance fixed as we move the poles apart and around the unit circle to vary the peak frequency respons We are interested in the low frequency effects that we have predicted to impact our RA estimates. Thus we w vary the peak frequency from near zero to about twice the frequency corresponding to the data span duration. T autocorrelation function corresponding to this noise process is given by Equation (26).

The attitude accuracy in RA and DEC in response to a moderately narrowband noise and to an extreme narrowband noise is shown in Figure 16. The extremely narrowband noise may be thought of practically as a si wave of fixed frequency and unit amplitude but random phase. As predicted by the discussion in the previo subsection, the frequencies near 2/3 of the data span frequency have the worst effect on RA accuracy. The DE accuracy, on the other hand, improves as the dominant frequencies get higher.

The accuracies that would result from the optimal data weighting are included in Figure 16, illustrating that in th colored noise case the weighting can make a significant difference to the estimator accuracy.



(Upper curve for suboptimal/unweighted estimator; lower curve for optimal/weighted estimator)

Figure 16. Standard Deviation Uncertainty Versus Low Frequency Noise From 0 to 2 Cycles Per Data Span

EXPECTED EFFECTS FOR LONGER DATA SPANS

looking at how the basis vector autocorrelation and power spectal density change as the data span increases, it ɔossible to make some general predictions about the effects that may be expected from colored noise and biases longer data spans are used. Figure 17 shows the RA and DEC basis vectors and their autocorrelation functions selected lengths of data spans. The characteristic shapes seen in Figure 15 for the short span is still seen until ire than about half an orbit is accumulated. Thus RA remains most sensitive to noise periods of about 2/3 of data span and DEC remains most sensitive to random biases. As the data span gets beyond one orbit the ocorrelation functions for RA and DEC undergo a transition in their shapes so that for two or more orbits both similar: a cosine function shaped by a triangular window in amplitude. (In the limit of long spans, this ıstrates how the cosine wave is the autocorrelation for a signal with random phase.) The power spectral density ewise undergoes transition from DEC sensitive to the zero frequency and RA sensitive to just the two lowest n-zero frequencies in the discret transform, evolving to both RA and DEC sensitive primarily to the orbit quency.

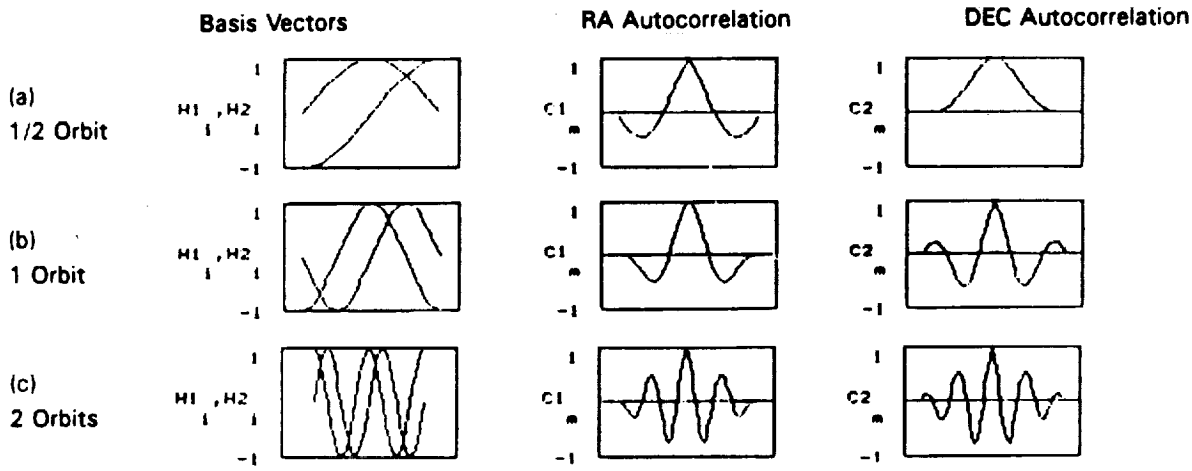


Figure 17. Basis Vectors and Their Autocorrelation Function for Longer Data Spans

us in multi-orbit data spans, neither RA nor DEC is sensitive to a random bias, and both are most sensitive to ise frequencies at orbital frequency. Since many physical phenomena occur at orbit frequencies (e.g., acecraft temperatures, orbit altitude, atmospheric drag, magnetic field changes, and science instrument erations), it is a useful to remember that any unmodeled or random aspects of their effects on sensor easurements are a potential source of noise with frequency content to which attitude solutions are most sensitive.

ıe effects of relatively short term correlations, on the other hand, can be shown to remain quite constant in ms of a correction factor as the data span increases. To understand this, keep in mind that the time scales are creasing in Figures 17 (a) through (c), and an autocorrelation function representing short term correlations stays queezed with the time scale) inside the main central peak which is always found. Thus the correction factor ɔm the projection can be expected to converge quickly.

BRIEF DISCUSSION OF GENERAL RESULTS

ıe results described above can be generalized for what we can call "well behaved cases:" those where the basis ctor frequency content is low relative to the data sampling frequency. This would apply, for example, to any t of orthogonal low order polynomials. An ideal set of basis vectors from the frequency analysis standpoint is a ıite Fourier series; then the basis vector power spectral densities are spikes at each of the lowest frequencies in e discrete transform. Polynomials would show a similar behavior with each term of higher order showing a

For these cases we can expect that the effects of short period correlations can be accomodated by a correctio factor in predicting the estimator accuracy, and unweighted least squares will perform almost as well as optima weighting. Cases where short term correlations can still impact the accuracy significantly will occur, fo example, in cases where the discrimination of two basis vectors relies heavily on a relatively few observatio close in time.

7. CONCLUSIONS

Techniques for analyzing the effects of colored noise on unweighted least squares accuracy have been explored, an an illuminating interpretation of the effects has been presented. These techniques were applied to some simple bu representative sample cases to show the colored noise impacts. More work remains to be done to apply thes techniques to additional and more complex cases, but nevertheless certain important conclusions may be draw from the general analysis and the cases already explored.

1. If a model for the actual noise correlations is available, the actual accuracy of the unweighted estimator ca be evaluated directly (without requiring a matrix inverse). This is recommended.

2. In certain commonly encountered well behaved cases (moderately lowpass noise and very low frequenc content in the basis functions), the effects of relatively short period correlations can be accommodated by simple correction factor to the white noise accuracy. This can be applied as a correction to the assume white noise standard deviation.

3. In these well behaved cases the optimally weighted estimator does not perform a lot better than th unweighted estimator. In this sense the unweighted least squares can be justified with colored noise, but th proper formula should be used to compute the expected uncertainty of the parameter estimates.

4. In general noise frequencies that are concentrated near the frequencies of the basis functions have th greatest impact on the accuracy of the corresponding parameter, as might be expected. This is quantifie mathematically in the frequency domain projection interpretation of the white noise correction factor.

5. Noise frequencies with corresponding periods of about 2/3 the data span length have the worst impac when an approximately linear (constant slope) term is being fit to the data.

6. Shorter data spans can be expected to be more sensitive to noise correlations particularly because correlations with time constants on the order of the data span are more likely.

7. The techniques described here can also be used to consider the effects of random biases on the solution accuracy.

Much further work can be done to extend the above results more generally and also more specifically to relevant applications. The author believes there is yet more to be explored in the relationship between spectral analysis and least squares solution accuracy. Since noise spectral content is shown to have a notable effect on the predicted accuracy of data fits, a key to improved knowledge of actual accuracies is improved knowledge of the spectral content of sensor noise.

REFERENCES

1. Bilanow, S. and M. Phenneger, *The Response of the SEASAT and MAGSAT Infrared Horizon Scanners to Cold Clouds*, Proceedings of the Fifth Annual Flight Mechanics/Estimation Theory Symposium, Goddard Space Flight Center, Greenbelt, 1980.

2. Bilanow, Stephen, Lily C. Chen, Davis W. Minor, and John P. Stanley, *LANDSAT-4 Horizon Scanner Performance Evaluation*, GSC-TR8401, General Sciences Corporation, prepared for Goddard Space Flight Center, 1984.

3. Markley, F. Landis, Ed, Seidewitz, and Mark Nicholson, *A General Model for Attitude Determination Error Analysis*, Proceedings of the Flight Mechanics/Estimation Theory Symposium, Goddard Space Flight Center, 1988.

4. Nicholson, M, F. Markley, and E. Seidewitz, *Attitude Determination Error Analysis System (ADEAS) Mathematical Specifications Document*, CSC/TM-88/6001, Computer Sciences Corporation, 1987.

5. Johnston, J, *Econometric Methods*, 2nd Edition, McGraw-Hill Kogakusha, Ltd. Tokyo, 1963.

5. Sage, Andrew P. and James L. Melsa, *Estimation Theory with Applications to Communications and Control*, Information and Control Sciences Center, Institute of Technology, Southern Methodist University, McGraw-Hill Book Company, 1971.

7. Bryson, A.E., Jr. and D.E. Johansen, *Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise*, 1964.

8. Bryson, A.E., Jr. and L.J. Henrikson, *Estimation Using Sampled Data Containing Sequentially Correlated Noise* J. Spacecraft, Vol. 5, No. 6, 1968.

9. Stear, Edwin B. and Allen R. Stubberud, *Optimal filtering for Gauss-Markov noise*, Int. J. Control, Vol. 8, No. 2, pp. 123-130, 1968.

10. Mehra, Raman K., and Arthur E. Bryson, Jr., *Linear Smoothing Using Measurements Containing Correlated Noise with an Application to Inertial Navigation*, IEEE Transactions on Automatic Control, Vol. AC-13, No. 5, 1968.

11. Johnson, Donald J., *Application of a Colored Noise Kalman Filter to a Radio-Guided Ascent Mission*, J. Spacecraft, Vol. 7, No. 3, 1970.

12. Kailath, Thomas and Roger A. Geesey, *An Innovations Approach to Least-Squares Estimation--Part V: Innovations Representations and Recursive Estimation in Colored Noise*, IEEE Transactions on Automatic Control, 1973.

13. Nahi, Nasser E., *Estimation Theory and Applications*, John Wiley & Sons, Inc., 1969.

14. Gelb, Arthur, ed., *Applied Optimal Estimation*, The M.I.T. Press, 1974.

15. Bryson, Arthur E., Jr. and Yu-Chi Ho, *Applied Optimal Control Optimization, Estimation, and Control*, Hemisphere Publishing Corp., 1975.

16. Oppenheim, Alan V. and Ronald W. Schafer, *Digital Signal Processing*, Prentice-Hall, Inc., 1975.

17. Tretter, Steven A., *Introduction to Discrete-Time Signal Processing*, John Wiley & Sons, 1976.

18. Box, George, E.P. and Gwilym M. Jenkins, *Time Series Analysis forecasting and control*, Holden-Day 1976.

19. Brockwell, Peter J. and Richard A. Davis, *Time Series: Theory and Methods* Springer-Verlag, 1987.

20. Kay, Steven M., *Efficient Generation of Colored Noise*, Proceedings of the IEEE, Vol. 69, No. 4, 1981.